REMARKS

Claims 13 through 25 are pending in this application. The Examiner rejected claims 13, 14, 17-18, 20, 22 and 24-25 under 35 U.S.C. § 103 as unpatentable over U. S. Patent No. 5,850,629 to Holm et al. in view of U. S. Patent No. 5, 557,661 to Yokoyama. The Examiner also rejected claims 19-21 and 23 under 35 U.S.C. § 103 over Holm et al. in view of Yokoyama as applied to claim 15, and further in view of U. S. Patent No. 5,313,522 to Slager. The Examiner objected to Figure 1 since it should be labeled as "Prior Art" and objected to the ABSTRACT as not being in proper format. The Examiner further objected to claim 14 since certain terms therein were unclear.

The Applicants have carefully reviewed and considered the Office Action. Attached please find a replacement sheet having Figure 1 labeled as "Prior Art" thereon. Applicants have amended the ABSTRACT in accordance with the Examiner's statement for clarity and brevity. Claims 13 and 14 have been amended to distinguish the invention over the art of record and to overcome the Examiner's objection to claim 14. In particular, regarding claim 13 the information relating to the individual property comprises gender, age, accent, pronunciation and a speech rate of synthesized speech. In accordance with the present invention, these individual properties are added to synthesized speech utilizing this additional information. Regarding claim 14, if information relating to texts and prosody parameter values exist in multimedia input information, the prosody parameter values in multimedia input information are utilized for generating the synthesized speech without the need for a procedure that generates the prosody symbols in a language processor and the prosody parameter values are calculated by a prosody processor. If text only information exists in multimedia input information, the values of prosody control parameters are calculated with the existing prosody processor of a text-to-text

speech conversion system, and these parameter values are utilized for generating the synthesized speech.

In light of the foregoing and further in view of the following, reconsideration and withdrawal of the rejection of the claims, and the objections to the claims and specification are earnestly solicited.

The Invention

Applicants have invented text-to-text speech conversion systems (TTS) for interlocking with multimedia and methods for organizing input data of the TTS which can enhance the production of natural synthesized speech. The invention accomplishes synchronization of multimedia with the TTS by defining additional prosody information required to interlock the TTS with the mulitmedia and interfacing this additional information with the TTS for use in the production of synthesized speech.

Traversal of the Rejections

The Examiner rejected claims 13, 14, 15-18, 20, 22, 24-25 under 35 U.S.C. § 103 over Holm et al. in view of Yokoyama. The Examiner found that with regard to claims 13 and 14, Holm et al. teach all of the elements of these claims except a picture output apparatus (claim 13) and synchronizing the prosody information with a moving picture and picture information (claims 13 and 14). The Examiner further found that Yokoyama teaches a system for coding and decoding moving pictures based on the result of speech analysis for interlocking media and so it would have been obvious to those of ordinary skill in the art to combine the teachings of Holm et al. with the teachings of Yokoyama to obtain the claimed invention. Applicants respectfully

traverse this rejection.

The methods and systems disclosed in Holm et al. relate to methods for setting up the environment of a synthesizer using a graphic user interface, and transmitting to the synthesizer by selecting the arbitrary part of text files to output the synthesized speech. Thus Holm et al. do not teach or suggest a structure for processing the received information comprising individual properties, prosody and lip shapes by text. Moreover, if the additional prosody information is input from an exterior source such as occurs in the present invention, this information can not be used in the methods and devices of Holm et al. Thus, the methods of the present invention can not be practiced by systems disclosed by Holm et al.

Furthermore, Holm et al. relates only to a user interface controller for a TTS synthesizer. The user can output a part or a series of text including synthesized speech without distorting the rhythm as if played from recording media such as a tape. Thus, the systems and methods of Holm et al. only output the synthesized speech by transmitting this text to a synthesizer after selecting the operating environment and the text to be synthesized in the synthesizer utilizing a graphic user interface. Therefore, the Holm et al. reference has nothing to do with accomplishing synchronization between multimedia and TTS, embodying the individual properties, and enhancing the natural production of synthesized speech, and does not teach or suggest a method for including information of lip shapes, individual property information and information relating to synchronization with moving picture transmission data.

In contrast, the present invention as recited in amended claims 13 and 14 covers a method for organizing input data of the TTS for accomplishing synchronization between multimedia and TTS, embodying the individual properties, and enhancing the production of natural synthesized speech. This is neither taught nor suggested by the art of record and

therefore, the obviousness rejection of claims 13 and 15 can not stand.

Yokoyama teaches coding and decoding speech and images wherein a small number of moving pictures are requested to be transmitted and displayed for some phonemes of dominant lip shape variations, making use of a telecommunication line of a low bit rate. Thus Yokoyama provides a speech and moving picture coding and decoding system wherein pictures which are close to real pictures and look more natural are displayed by a simple method without the necessity of complicated processing for picture analysis or picture synthesis.

However, Yokoyama has nothing to do with accomplishing synchronization between multimedia and TTS, embodying individual properties, and enhancing synthesized speech. Moreover, Yokoyama neither teaches nor suggests a method for including information of lip shapes, information of individual property, and information of synchronization with the moving picture on transmitting data.

Additionally, Yokoyama relates only to a method for transmitting/receiving/playing the natural moving picture and speech signal on the basis of phonemes which are estimated from analyzing the natural voice. Similarly, the methods and devices disclosed in Yokoyama are wholly unrelated to interlocking between TTS and multimedia. The methods and systems of Yokoyama can not guarantee the succession of a moving picture as is accomplished according to the present invention. Therefore, Holm et al. in combination with Yokoyama neither teach nor suggest the invention of amended claims 13-15, 17-18, 20, 22 and 24-25 which require interlocking and synchronization. Reconsideration and withdrawal of the rejection of these claims under 25 U.S.C. § 103 are therefore earnestly solicited.

The Examiner also rejected claims 19-21 and 23 under 35 U.S.C. § 103 over

Holm et al. in view of Yokoyama and further in view of Slager, stating that Slager additionally teaches generating from a speech signal a moving visual lip image from which one can deduce the speech content of the signal. Therefore, the Examiner found that it would have been obvious to combine Holm et al. with Yokoyama and Sager to obtain the invention of claims 19-21 and 23. Applicants also respectfully traverse this rejection.

The object of the system and methods disclosed in Slager is to provide a device which can increase the ability of a hearing-impaired person to comprehend a telephone conversation, especially in a high background noise situation. This object is met by providing an apparatus which includes a circuit, an arrangement for coupling the circuit to a telephone line, and a display arrangement coupled to the circuit. The circuit includes an arrangement for breaking a received audio speech signal into a series of successive phonemes and an arrangement responsive to the series of phonemes for displaying on the monitor an image of a human lip forming a succession of lip shapes respectively corresponding to respective phonemes in the series. Also, Slager is directed to a method for forming human lips utilizing a mapping table in a receiver after recognizing the phoneme from the input speech.

Slager is in no way related to a method for including information of lip shapes in transmitting data, nor with accomplishing synchronization between multimedia and TTS to enhance speech. However, the present invention as recited in claims 19-21 and 23 relates to a method for organizing input data of the TTS, for accomplishing synchronization between multimedia and TTS, embodying the individual properties, and enhancing the natural production of synthesized speech.

In accordance with the present invention, speech is synthesized on the basis of lip shapes which are assumed from a moving picture or made out by manual input data of a TTS

organized for establishing the synchronization of the synthesized speech. This is accomplished by calculating the duration time of each phoneme from the moving picture in order to compare time and align lip shapes which are assumed from a series of phonemes of the TTS. Moreover, the present invention accomplishes synchronization between the moving picture and the synthesized speech generated from the TTS using information relating to output time of synthesized speech by text.

Consequently, the present invention differs significantly from the systems disclosed by Holm et al., Yokoyama and Slager since none of these prior art references teach or even remotely suggest interlocking and synchronizing between multimedia and a TTS as is accomplished according to the present invention. It is respectfully submitted that Slager adds nothing to Holm et al. and Yokoyama that could render claims 19-21 and 23 obvious. Reconsideration and withdrawal of the rejection of these claims as obvious in light of these references is therefore earnestly solicited.

CONCLUSION

Applicants have invented new and unobvious text-to-text speech conversion systems for interlocking with multimedia which are neither taught nor suggested by the art of record. The art cited by the Examiner but not used to reject the claim has been considered and it is submitted that it neither anticipates nor renders obvious the claimed invention. As this application thus appears in a condition of allowance, a prompt Notice of Allowance is earnestly solicited.

It is believed that no additional fees or charges are required at this time. However, if any additional fees or charges are required at this time in connection with the application, they may be charged to our Patent and Trademark Office Deposit Account No. 03-2412.

Respectfully submitted,

COHEN, PONTANI, LIEBERMAN & PAVANE

By _____

Jeffrey M. Navon
Reg. No. 32,711
551 Fifth Avenue, Suite 1210
New York, New York 10176
(212) 687-2770

Dated: October 12, 1999